BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e137853
doi: 10.3897/biss.8.137853

OPEN ACCESS

Conference Abstract

GBIF's Vocabulary Server: A Tool to Create, Manage and Apply Controlled Vocabularies for Biodiversity

Cecilie Svenningsen‡, Marcos L Gonzalez‡, Tim Robertson‡

‡ Global Biodiversity Information Facility, Copenhagen, Denmark

Corresponding author: Cecilie Svenningsen (csvenningsen@gbif.org)

Received: 25 Sep 2024 | Published: 26 Sep 2024

Citation: Svenningsen C, Gonzalez M, Robertson T (2024) GBIF's Vocabulary Server: A Tool to Create, Manage and Apply Controlled Vocabularies for Biodiversity. Biodiversity Information Science and Standards 8: e137853. https://doi.org/10.3897/biss.8.137853

Abstract

Global Biodiversity Information Facility's (GBIF) global index of primary biodiversity data is based on contributions from more than 2,200 publishing institutions and over 100,000 datasets. Data originate from a broad variety of fields, from research to citizen science, from eDNA through specimen collections to observations and monitoring projects, across all taxonomic groups, and with a wide range of datasets and data types. Equally variable is the underlying motivation and focus for data collection and digitization. Even within commonly used domain concepts like Kingdom or OccurrenceStatus, the values provided by the original sources can vary greatly. This presents challenges for any system that aims to provide joint search access across all these different resources.

The vocabulary server*[1] in the GBIF Registry is a tool used to standardize selected fields during data interpretation, thereby increasing the searchability of records across datasets on GBIF.org and other GBIF-hosted websites. The server hosts controlled vocabularies for relevant terms from the Darwin Core standard*[2], Global Registry of Scientific Collections (GRSciColl)*[3], and specific fields used in GBIF. The vocabulary development is managed in a GitHub repository*[4], and both concept development and the mapping of verbatim values to concepts are supported by community participation and expert groups.

© Svenningsen C et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

During this talk, we will present some of the vocabularies implemented on GBIF.org and the work that is going into it. We will also go through the roadmap*[5] for future implementations and vocabularies.

We value input from the Society for the Preservation of Natural History Collections (SPNCH) and Biodiversity Information Standards (TDWG) communities on terms relevant to controlled vocabulary development in GBIF.

Keywords

data standards, data interoperability, data digitization, global data sharing, data searchability, community-driven vocabularies

Presenting author

Cecilie Svenningsen

Presented at

SPNHC-TDWG 2024

Conflicts of interest

The authors have declared that no competing interests exist.

Endnotes

*1 https://registry.gbif.org/vocabulary/search
*2 https://dwc.tdwg.org/
*3 https://scientific-collections.gbif.org/
*4 https://github.com/gbif/vocabulary
*5 https://gbif.link/vocabulary-server-roadmap